HAL manuscript No. 00399832

Transportation Distances on the Circle and Applications

Julie Delon · Julien Rabin · Yann Gousseau

Received: date / Accepted: date

Abstract This paper is devoted to the study of the Monge-Kantorovich theory of optimal mass transport and its applications, in the special case of *one-dimensional and circular distributions*. More precisely, we study the Monge-Kantorovich distances between discrete sets of points on the unit circle S^1, in the case where the ground distance between two points x and y is defined as $h(d(x,y))$, where d is the geodesic distance on the circle and h a convex and increasing function. We first prove that computing a Monge-Kantorovich distance between two given sets of pairwise different points boils down to cut the circle at a well chosen point and to compute the same distance on the real line. This result is then used to obtain a metric between 1D and circular discrete histograms, which can be computed in linear time. A particular case of this formula has already been used in [RDG09] for the matching of local features between images, involving circular histograms of gradient orientations. In this paper, other applications are investigated, in particular dealing with the hue component of color images. In a last part, a study is conducted to compare the advantages and drawbacks of transportation distances relying on convex or concave cost functions, and of the classical L^1 distance.

Keywords Optimal mass transportation theory · Earth Mover's Distance · Circular histograms · Image retrieval

PACS PACS code1 · PACS code2 · more

Mathematics Subject Classification (2000) MSC code1 · MSC code2 · more

1 Introduction

The theory of optimal transportation was first introduced by Monge [Mon81] in its *Mémoire sur la théorie des déblais et des remblais* (1781) and rediscovered by Kantorovich [Kan42] in the late '30s. The Monge-Kantorovich problem can be described in the following way. Given two probability distributions f and g on X and c a nonnegative measurable cost function on $X \times X$, the aim is to find the optimal transportation cost

$$\mathrm{MK}_c(f,g) := \inf_{\pi \in \Pi(f,g)} \iint_{x,y} c(x,y)\, d\pi(x,y) \tag{1}$$

This work has been realized under grant BLAN07-2_183172.

Julie Delon and Yann Gousseau
CNRS LTCI Télécom ParisTech, 46 rue Barrault, 75634 Paris Cedex 13
Tel.: +33-145-817-073
Fax: +33-145-813-794
E-mail: julie.delon@telecom-paristech.fr and yann.gousseau@telecom-paristech.fr

Julien Rabin
CNRS CÉRÉMADE Université Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16
Tel.: +33-144-054-923
Fax: +33-144-054-599
E-mail: julien.rabin@ceremade.dauphine.fr

where $\Pi(f,g)$ is the set of probability measures on $X \times X$ with marginals f and g (such measures are called transportation plans). The existence, uniqueness and behavior of optimal transportation plans has been thoroughly studied in the last decades[Vil03, ACB+03, McC99, Vil08, McC95, GM96].

This framework is nowadays widely used in many fields of research, such as cosmology [FMMS02], meteorology [Cul06], fluid mechanics or electromagnetic (see [ACB+03] for a complete review).

The use of the Monge-Kantorovich framework in image processing and computer vision has been popularized by Rubner *et al.* [RTG00] for image retrieval and texture classification with the introduction of the so-called Earth Mover's Distance (EMD). Although the definition of the EMD is slightly different from the original Monge-Kantorovich formulation, these are equivalent when considering distributions having the same total weight. In the following years and up to now, a large body of works has relied on the use of such distances for image retrieval, see e.g. [GDR00, LCL04, LZLM05, ZWG06, HGS08, PW09]. This extensive use of transportation distances is largely due to their robustness when comparing histograms or discrete distances. For the same reason, these distances are also successfully used to compare local features between images, see [LO07, PW08, RDG08, RDG09, PW09]. Other uses of transportation distances for images include: image registration [HZTA04], image morphing [ZYHT07] or junction detection [RT01].

The strongest limitation of transportation distances is their computational cost. Standard approaches quickly become intractable when dealing with a large amount of data in dimensions more than two. Indeed, the simplex algorithm, interior point methods or the Hungarian algorithm all have a complexity of at least $0(N^3)$ (N being the size of the data, either the number of samples or the number of histogram bins). Therefore, several works have proposed to speed up the computation or the approximation of optimal transport, in particular in the field of image processing, where the amount of data is often massive, see [IT03, GD04, LO07, SJ08]. One particular case in which the computation is elementary and fast is the case of *one-dimensional histograms*, for which it is well known that optimal transport, in the case of a *convex cost function*, reduces to the pointwise difference between cumulative distribution functions [Vil03]. A question that arises is then the possibility to perform such simple and efficient computations in the case of *circular* histograms, i.e. histograms in which the first and last bins are neighbors.

Indeed, circular histograms are especially important in image processing and computer vision. First, the local geometry is often efficiently coded by the distribution of gradient orientations. Such representations offer the advantage of being robust to various perturbations, including noise and illumination changes. This is in particular the case for the well known SIFT [Low04] descriptor and its numerous variants. In such a situation, the comparison of local features reduces to the comparison of one-dimensional *circular* histograms. Other local features involving circular histograms include the so-called Shape Context [BMP02]. Second, the color content of an image is often efficiently accounted for by its *hue*, in color spaces such as HSV or LCH. In such cases again, information is coded in the form of circular one-dimensional histograms. Several works in the field of computer vision have explicitly addressed the use of transportation distances in the case of circular histograms, either using thresholded concave cost functions [PW08, PW09] or L^1 cost functions [RDG08, RDG09].

The goal of this paper is first to give a general formulation of transportation distances when the cost function is a convex function of the Euclidean distance on the circle. This formulation gives a practical way to compute distances in linear time in this case. Second, we provide various experiments of image manipulation or retrieval for which the interest of circular transportation distances is shown. Eventually, we conclude with a discussion (that actually applies to both circular and non-circular cases) on the respective interest of transportation distances with either convex or concave cost functions when compared to classical bin-to-bin distances. It is shown that the choice between these three family of distances should essentially be driven by the type of perturbation the histograms are likely to suffer from.

Outline The paper is organized as follows. In Section 2, the optimal transportation flow of the Monge-Kantorovich problem is investigated in the circular case. The definition of this problem being recalled, a new formula is introduced and a sketch of the proof is proposed (details of the proof are provided in appendix A). In section 3, several applications are studied to show the interest of such a metric for image processing and computer vision. First, an application to hue transfer between images is proposed in § 3.1 as a result of an optimal transportation flow on the circle. Then, in § 3.2, some applications of histogram comparison for image retrieval are proposed. Eventually, the discussion about the robustness and the limitations of Monge-Kantorovich distances in the framework of histogram comparison is given in Section 4.

2 The Monge-Kantorovich transportation problem on the circle

In this section, we present some results on the Monge-Kantorovich distances between two circular histograms. In particular, we give an analytic formulation of these distances when the ground cost between points on the circle can be written as an increasing and convex function of the Euclidean distance along the circle.

2.1 Definitions

Consider two discrete and positive distributions

$$f = \sum_{i=1}^{N} f[i]\delta_{x_i} \quad \text{and} \quad g = \sum_{i=1}^{M} g[j]\delta_{y_j}, \tag{2}$$

where $\{x_1, \dots x_N\}$ and $\{y_1, \dots y_M\}$ are two sets of points on a subset Ω of $\mathbb{R}^K$. Assume that these distributions are normalized in the sense that $\sum_{i=1}^{N} f[i] = \sum_{j=}^{M} g[j] = 1$. Let $c : \Omega \times \Omega \mapsto \mathbb{R}^+$ be a nonnegative cost function (called *ground cost*), the quantity

$$\mathrm{MK}_c(f, g) := \min_{(\alpha_{i,j}) \in \mathcal{M}} \sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_{i,j} c(x_i, y_j), \text{ with} \tag{3}$$

$$\mathcal{M} = \{(\alpha_{i,j}) \in \mathbb{R}^N \times \mathbb{R}^M;\ \alpha_{i,j} \geq 0,\ \sum_i \alpha_{i,j} = g[j],\ \sum_j \alpha_{i,j} = f[i]\}, \tag{4}$$

is called the *optimal transportation cost* between f and g for the ground cost c. Matrices $(\alpha_{i,j})$ in $\mathcal{M}$ are called *transport plans* between f and g. If $(\alpha_{i,j})$ is optimal for (3), we say that $(\alpha_{i,j})$ is an *optimal transport plan*.

Let d be a distance on Ω and assume that the ground cost can be written $c(x, y) = d(x, y)^\lambda$, with the convention $d(x, y)^0 = \mathbf{1}_{x \neq y}$. It can be shown [Vil03] that

- when $\lambda \in [0, 1[$, MK_c is a distance between probability distributions ;
- when $\lambda \in [1, \infty[$, $\mathrm{MK}_\lambda(f, g) := (\mathrm{MK}_c(f, g))^{\frac{1}{\lambda}}$ is also a distance between probability distributions.

These distances are called *Monge-Kantorovich distances*, or Wasserstein distances. For $\lambda = 1$, MK_1 is also known as the Kantorovich-Rubinstein distance, or in computer vision as the *Earth Mover's Distance* (EMD), as introduced by Rubner in [RTG00].

Computing optimal transportation costs is generally not an easy task. The main exception is the case of the real line: if $\Omega = \mathbb{R}$, and if the cost c is a convex and increasing function of the euclidean distance $|x - y|$, then the optimal transport plan between f and g is the monotone rearrangement of f onto g, which sends the mass starting from the left. This result is usually false if c is not a convex function of the euclidean distance on the line.

In the following, we take interest in the case where Ω is a circle S^1 of perimeter 1, and where c is an increasing function of the geodesic distance d along the circle. In particular, we will see that the previous result on the line can be generalized in this case.

2.2 Optimal transportation for convex functions of the distance

The main result of this section is an analytic formulation of the optimal transportation cost between the discrete distributions f and g on the circle S^1 when the ground cost c can be written $c(x, y) = h(d(x, y))$, with $h : \mathbb{R} \to \mathbb{R}^+$ an increasing and convex function and d the geodesic distance along the circle. In the following, we use the same notations for points on the circle and their coordinates along the circle, regarded as variables taking their values on the reduced interval $[0, 1[$ $(modulo\, 1)$. It follows that d can be written

$$d(x, y) = \min(|x - y|, 1 - |x - y|). \tag{5}$$

The distributions f and g on S^1 can be seen equivalently as periodic distributions of period 1 on $\mathbb{R}$.

Let us define the cumulative distribution function of f on $[0,1[$ as

$$\forall y \in [0,1[, \quad F(y) = \sum_{i=1}^{N} f[i] \cdot \mathbb{1}_{\{x_i \in [0,y[\}} \ . \tag{6}$$

F is increasing and left continuous, and can be extended on the whole real line with the convention $F(y+1) = F(y) + 1$. This boils down to consider f as a periodic distribution on $\mathbb{R}$. We define also the pseudo-inverse of F as $F^{-1}(y) = \inf\{t;\ F(t) > y\}$. The interest of these definitions lies in the next result.

2.2.1 An analytic formulation of optimal transportation on the circle

Theorem 1 *Assume that d is given by Equation (5) and that the ground cost c can be written $c(x,y) = h(d(x,y))$, with $h : \mathbb{R} \to \mathbb{R}^+$ an increasing and convex function. Let f and g be two discrete probability distributions on the circle, with cumulative distribution functions F and G, and let G^α denote the function $G - \alpha$. Then*

$$\mathrm{MK}_c(f,g) = \inf_{\alpha \in \mathbb{R}} \int_0^1 h(|F^{-1} - (G^\alpha)^{-1}|) \ . \tag{7}$$

Idea of the proof This result is a generalization of the real line case, where it is well known [Vil03] that the global transportation cost between two probability distributions f and g can be written

$$\mathrm{MK}_c(f,g) = \int_0^1 h(|F^{-1} - G^{-1}|) \ . \tag{8}$$

A proof of Equation (7) in a continuous setting has been proposed very recently in [DSS10], where it is shown that this equation holds for any couple of probability distributions. However, this proof involves some complex notions of measure theory which are not needed in the discrete setting. For the sake of completeness and simplicity, we provide in Appendix A a simpler proof of these theorem in the case of discrete distributions. The proof first focus on the case where f and g can be written as sums $f = \frac{1}{P}\sum_{k=1}^{P} \delta_{x_k}$, and $g = \frac{1}{P}\sum_{k=1}^{P} \delta_{y_k}$, where $\{x_1, \dots x_P\}$ and $\{y_1, \dots y_P\}$ are discrete sets of points on the unit circle S^1. When the points are all pairwise different, we show that the circle can always be "cut" at some point, such that computing the optimal transport between f and g boils down to compute an optimal transport between two distributions on the real line (see Figure 1). This result is proven first for strictly convex functions h and for any optimal transport plan, then for any convex function h and a well chosen optimal plan. Once the problem has been reduced to the real line, Formula (7) follows from the fact that the optimal transport on $\mathbb{R}$ is given by the ordering of the points. The generalization of this formula to any kind of discrete distribution results from the continuity of the global transport cost $\mathrm{MK}_c(f,g)$ in the values of the masses and their positions on the circle.

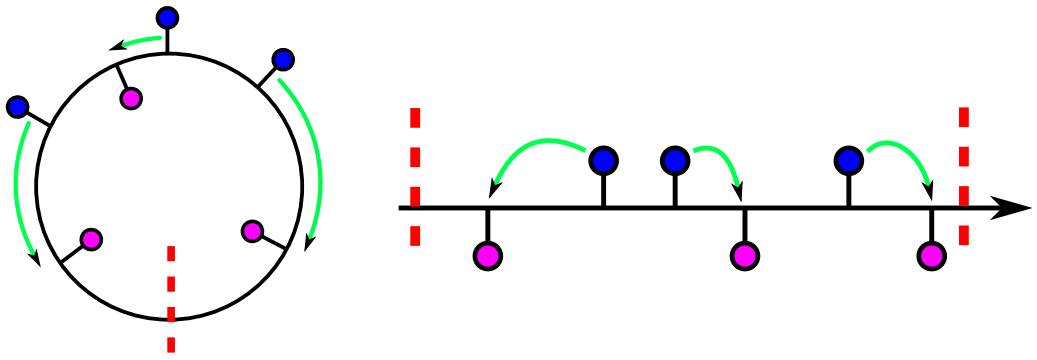


Fig. 1 When the distributions are sums of unitary different masses, and when the ground cost is a nonnegative, increasing and convex function of the distance along the circle, there is a (non-necessarily unique) "cut" on the circle such that the optimal transportation on S^1 boil down to optimal transportation on the real line.

In practice, Formula (7) can be computed for any h at a precision ε with a complexity in $O((N+M)\log\frac{1}{\varepsilon})$ [DSS10], where N and M are the number of points in the distributions f and g (*i.e.* the number of masses in the distributions).

2.2.2 The case $c(x,y) = d(x,y)$

If h is a power function $x \mapsto x^\lambda$, with $\lambda \geq 1$, Theorem 1 gives us a way to compute Monge-Kantorovich distances between f and g:

$$\mathrm{MK}_\lambda(f,g) = \left(\inf_{\alpha\in\mathbb{R}} \int_0^1 |F^{-1} - (G^\alpha)^{-1}|^\lambda \right)^{\frac{1}{\lambda}}. \tag{9}$$

In the case $\lambda = 1$ (*i.e.* when the ground cost c is the distance d along the circle), this result can be rewritten

$$\mathrm{MK}_1(f,g) = \inf_{\alpha\in\mathbb{R}} \int_0^1 |F - G - \alpha|. \tag{10}$$

Observe that an alternative proof of Equation (10) was proposed by Werman et al. in [WPMK86] for distributions written as sums of unitary masses. A similar result is shown in [CM95] for the Kantorovich-Rubinstein problem, which is known to be equivalent (see [Vil03], chapter 1) to the Monge-Kantorovich problem when the cost $c(x,y)$ is a distance, which is true for $\lambda = 1$ (but false for $\lambda > 1$).

In practice, since $F - G$ is piecewise constant for discrete distributions, the infimum of Equation (10) can be computed easily by computing the weighted median of the (finite number of) values $F(t) - G(t)$ when $t \in [0,1[$, the weights being the lengths of the intervals on which $F - G$ is constant. In practice, this yields a $O(N)$ exact algorithm to compare two normalized distributions of N masses [CM98, Gur90].

2.2.3 Discrete histograms

Most applications deal with discrete histograms, *i.e.* discrete distributions living on a uniform grid of N bins. In the case of histograms on the real line, for the cost $c(i,j) = |i-j|$, Formula (8) becomes

$$\mathrm{EMD}(f,g) = \mathrm{MK}_1(f,g) = \|F - G\|_1, \tag{11}$$

where we denote by $\|.\|_1$ the discrete L^1 norm, by F and G the cumulative histograms of f and g, and where EMD is the acronym for Earth Mover's Distance [RTG00]. An illustration is given in Figure 2.

In the case of circular histograms, bins 0 and $N-1$ are neighbors. If the cost c is $c(i,j) = \min(|i-j|, N-|i-j|)$ along the circle, Formula (10) can be rewritten

$$\textsc{cemd}(f,g) := \mathrm{MK}_1(f,g) = \inf_\alpha \sum_{i=0}^{N-1} |F[i] - G[i] - \alpha| = ||F - G - \mu||_1, \tag{12}$$

where μ is the median of the set of values $\{F[i] - G[i], 0 \leq i \leq N-1\}$, and where CEMD is the acronym for Circular Earth Mover's Distance, as defined in [RDG09]. Indeed, it is easily checked that the distance defined by Formula (12) is equivalent to the distance introduced in [RDG09], that is

$$\textsc{cemd}(f,g) = \min_{k\in\{0,\dots,N-1\}} \|F_k - G_k\|_1, \tag{13}$$

where $F_k[i]$ is defined as $F[i] - F[k]$ if $i \in \{k, \dots N-1\}$ and $F[i] - F[k] + 1$ if $i \in \{0, \dots k-1\}$ (the definition being similar for G_k by replacing f by g). In other words, the distance $\mathrm{MK}_1(f,g)$ is the minimum in k of the L^1 distance between F_k and G_k, the cumulative histograms of f and g starting at the k^{th} quantization bin.

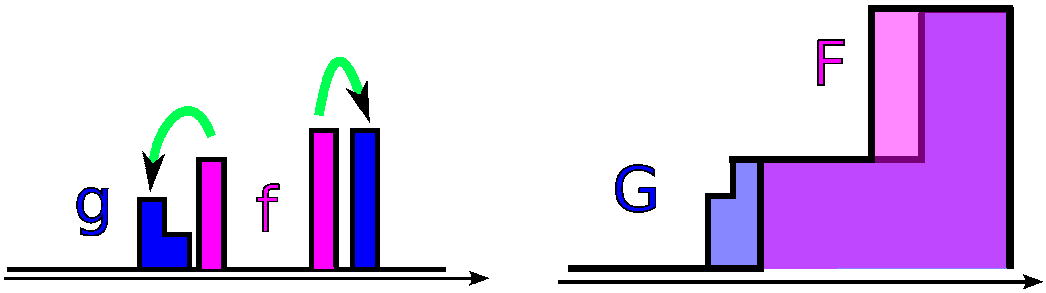


Fig. 2 Using $c(x,y) = |x-y|$ on the real line, the optimal transportation plan between two discrete histograms f and g is the L^1 distance of the difference between the cumulative histograms F and G (Formula (11)).

2.3 Optimal transportation for concave functions of the distance

In practice, it may be useful to choose the ground cost c as a nonnegative, concave and increasing function h of the ground distance d. For instance, for the task of image retrieval, several authors [RTG00, HGS08, RT01]) claim that optimal results can be achieved with a function

$$h(t) = 1 - e^{-\frac{t}{\tau}}. \tag{14}$$

Notice that if h is increasing, concave and such that $h(0) = 0$, it is easy to show that $h(d)$ is also a distance, and thus MK_c is also a distance between probability distributions. Another property of concave costs is that they do not move the mass which is shared by the distributions [Vil03]: if f and g are histograms, the problem is reduced to the comparison of $(f-g).\mathbf{1}_{f-g\geq 0}$ with $(g-f).\mathbf{1}_{f-g<0}$, which have disjoint supports.

However, in the case of such concave functions h, Theorem 1 does not apply, and there is no general and fast algorithm to compute corresponding optimal transportation costs, either on the real line or on the circle. In most cases, we are reduced to use linear programming, *i.e.* simplex or interior point algorithms, which are known to have at best a $O(N^3 \log N)$ complexity to compare two histograms on N bins [BDM09]. We describe in the following some special cases of concave function h for which this complexity can be reduced.

2.3.1 L^1 as a Monge-Kantorovich distance

If the distributions f and g are discrete histograms on N bins, and if $h(t) = \mathbf{1}_{t\neq 0}$, then the Monge-Kantorovich distance between f and g is [Vil03]

$$\mathrm{MK}_{\mathbf{1}_{d(x,y)\neq 0}}(f,g) = \frac{1}{2}\sum_{i=1}^{N} |f[i]-g[i]| = \|f-g\|_1. \tag{15}$$

In other words, the L^1 distance between two normalized histograms f and g is a Monge-Kantorovich distance for the concave function $\mathbf{1}_{t\neq 0}$.

2.3.2 Thresholded distances

In [PW08, PW09], Pele and Werman consider thresholded ground distances, using $h(t) = \min(t,T)$, with T a given threshold. Up to a multiplicative factor, this function h can almost be seen as a discrete version of (14), where τ is chosen proportional to T. They show that in this case, the computation of the optimal cost can be solved by a "min-cost-flow algorithm", whose complexity is smaller than classical linear programming algorithms. In their experiments, they use $T = 2$ for comparing histograms on N bins, which means that the cost $c(i,j)$ can take only three values: 0 if $i = j$, 1 if i and j are neighbors, and 2 in other cases. Since all the shared mass remains in place, we know that if $(\alpha_{i,j})$ is the optimal

transport plan between f and g, then for a given i, $\sum_{j\neq i}\alpha_{i,j} = (f[i]-g[i])\mathbf{1}_{f[i]\geq g[i]}$, which implies in particular that

$$\begin{aligned}\mathrm{MK}_{\min(d,2)}(f,g) &= \sum_{i=1}^{N}((\sum_{j\neq i}2\alpha_{i,j}) - \alpha_{i,i+1} - \alpha_{i,i-1}) \\ &= 2\sum_{i=1}^{N}(f[i]-g[i])\mathbf{1}_{f[i]\geq g[i]} - \sum_{i=1}^{N}(\alpha_{i,i+1}+\alpha_{i,i-1}) \\ &= \|f-g\|_1 - \sum_{i=1}^{N}(\alpha_{i,i+1}+\alpha_{i,i-1}).\end{aligned}$$

Now, notice that $\alpha_{i,i+1}$ is different from 0 only if $f[i] \geq g[i]$ (otherwise, all the mass in i stay in place) and $f[i+1] < g[i+1]$ (otherwise the mass $g[i+1]$ is already "filled" by a part of $f[i+1]$). In other words, the only points where the quantities $\alpha_{i,i-1}$ or $\alpha_{i,i+1}$ are different from 0 are the points where the densities of f and g are crossing. It follows that in many cases, the thresholded distance $\mathrm{MK}_{\min(d,2)}$ is very close to L^1, in particular when N is large and when histograms are crossing at only a few places, as we will see in the experiments of Section 4. In order to allow larger ground displacements, the use of values of T larger than 2 is proposed in [PW09]. This is made at the price of a non-linear complexity and necessitates a compromise in the tuning of T (smaller values yield faster computations). We will come back on the use of such concave cost functions in Section 4.

In the following section, we illustrate the interest of circular Monge-Kantorovich distances for several applications.

3 Experimental study

This section is devoted to an experimental analysis of the previous optimal transportation framework for different computer vision tasks. We first consider the color transfer from one image to another as an application of the optimal transportation flow on the circle between hue histograms. In the following, two different applications of transportation distances for the comparison of histograms are studied: local feature comparison and image retrieval.

3.1 Hue transfer between color images

The aim of this paragraph is to use the optimal transportation framework to transfer a hue distribution from one image to another.

Contrast transfer First, let us recall that histogram equalization and more generally histogram specifications are merely particular cases of optimal transportation on the real line. Indeed, if $u : \Omega \mapsto \{0,\dots,L-1\}$ is a discrete image and h_u its gray level distribution, histogram specification consists in finding the optimal transport plan between h_u and a target discrete probability distribution h_t on $\{0,\dots,L-1\}$ (one speaks of histogram equalization when h_t is a constant distribution on $\{0,\dots,L-1\}$). If one considers a cost c equal to the euclidean distance on the line, then, as explained in Section 2.1, the solution of this problem consists in a monotone rearrangement. This rearrangement is obtained by applying the function $H_t^{-1}\circ H_u$ to u, where H_u (resp. H_t) is the cumulative distribution function of h_u (resp. h_t) and H_t^{-1} represent the pseudo-inverse of H_t (see definition in section 2.2, after Formula (6)). If u is a color image, such contrast adjustments can be applied to its "intensity" channel (*e.g.* the channel "Value" in the HSV representation).

Hue transfer Thanks to Formula (12) or (13) (Section 2.2.2), one can extend the previous framework to hue distributions, seen as circular distributions. Following Equation (12), the optimal mapping between the hue distribution h_u of an image u and the target hue distribution h_t is obtained as $(H_t-\alpha)^{-1}\circ H_u$, where α is the median of the values $\{H_u[i]-H_t[i]\}$. Figure 3 illustrate this transfer of hue on a pair of images. For a detailed survey on color transfer between images, we refer the reader to [PKD07].

Fig. 3 Hue transfer between images. *First row:* original images. *Second row:* the hue channel of each image has been modified by applying the circular optimal transportation flow between the hue channels (see text for details), while other channels ("saturation" and "value" in HSV representation) are kept unchanged.

3.2 Image and histogram comparison

In this section, we investigate the interest of Monge-Kantorovich distances for histogram comparison. More precisely, the distance considered in the following experiments is the MK_1 distance (given by Equation (12)) on the circle. Following [RDG09], we refer to it as CEMD (Circular Earth Mover's Distance). This distance is compared in particular to the L^1 distance, which can be described as being bin-to-bin, since it compares bins having the same index.

3.2.1 Local feature comparison

Many computer vision tasks rely on local features (object recognition, image retrieval, indexation and classification, image mosaicing, *etc*). Some of the most commonly used (and invariant) local features are the Shape Context [BMP02] and SIFT descriptors [Low04], which both encode periodic data: polar orientation for the former, and gradient orientation for the latter. For instance, a SIFT descriptor can be seen as a collection of M one-dimensional and circular-histograms, each one being constructed from a subpart of a localization grid in the image domain [Low04].

In [RDG09], it is demonstrated that the Circular Earth Mover's Distance (or MK_1 distance on the circle) is far more robust than classical bin-to-bin distances (L^1 and L^2 metric, χ^2 distance, etc) to compare SIFT descriptors. In particular, it is underlined that this cross-bin distance is more adapted to two kinds of perturbations

- **quantization** effects (also know as the "binning problem"), which result from the small number of samples used to build the discrete histograms of gradient orientations, but also from the localization grid used to extract histograms over the pixel grid;
- **shifts** in histograms, resulting from geometrical deformations in the image (*e.g.* perspective effects, which typically arise when an object is seen from different points of view).

Note that this result is consistent with the one presented in [PW08], where the metric used to compare SIFT descriptors is a modification of the Monge-Kantorovich distance with a truncated cost c. A more in depth analysis of the general robustness of Monge-Kantorovich distances for both quantization and shift effects is proposed in section 4.

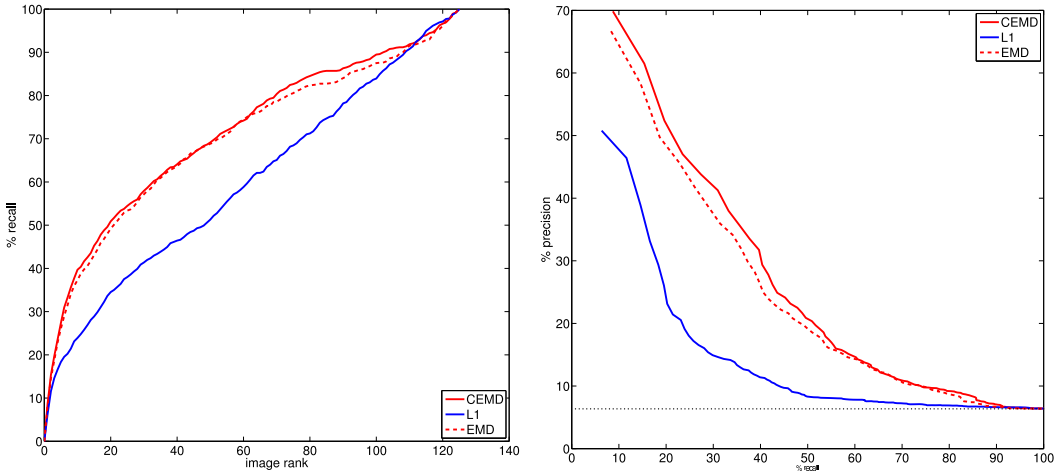
3.2.2 Three experiments on color image retrieval

For the task of color image retrieval, numerous studies have shown that the Earth Mover's Distance (defined in Section 2.1) often achieves better retrieval performances than bin-to-bin distances [RTG00, GDR00, RT01, Dvi02, LCL04, LZLM05, ZWG06, HGS08, PW09]. In order to illustrate the advantages of the Circular Earth Mover's Distance in the same context, we rely on hue distributions to perform image retrieval on a color image dataset. The dataset[1] contains 14 category of 9 pictures of the same object, with various camera settings (sensitivity, with or without flash, white balance reference, exposure time, *etc*). Nine pictures of the same category are shown as an example in Figure 4. Each of the $P = 14 \times 9 = 126$ images of the dataset is described by a hue (channel H of HSV representation) distribution, built on $N = 360$ bins. For a given dissimilarity measure D, the retrieval experiment consists in using an image of the dataset as a query and finding the r most similar images for D. For each value of r, we compute

- the **recall**, which is defined as the average, when the query spans the database, of the ratio between the number of correctly retrieved images among r and the size of the query class;
- the **precision**, which represents the average on the whole database of the rate of true positives among the r most-similar images.

The curves (r, recall) and (recall, precision) are drawn on Figure 5, for three different dissimilarity measures: CEMD (Equation (12)), non circular EMD (Equation (11)), and the L^1 distance.

Fig. 4 Example of a category of 9 pictures extracted from the image database used for image retrieval (results ate shown in Figure 5). These photographs represent the same scene under various illumination conditions and camera settings.



(a) Average Recall rate according to the rank

(b) Average precision-recall curve

Fig. 5 retrieval of a color image database. The performances curves are obtained using CEMD in red continuous line, L^1 distance in blue line and also EMD (non circular mass transportation) in red dashed line.

[1] the image dataset is available at the following address: http://perso.telecom-paristech/~rabin/database/

In this experiment, the results of CEMD (in red continuous line) clearly outperforms those of L^1 and EMD (which does not take into account the circularity of hue histograms). As in the experiments on local feature comparison, one can guess that the superiority of Monge-Kantorovich distances are due to their natural robustness to shifts in the distributions. However, observe that this time, in contrast with the application to local descriptors, a huge number of samples ($\sim 5.10^5$) is used to build each circular histogram, so that we avoid the aforementioned binning problem.

Two more color retrieval experiments In this paragraph, we aim at showing that two different classes of intraclass variability could arise when representing data by histograms, and that bin-to-bin and cross-bin distances behaves very differently according to these perturbations. This fact will then be discussed in more detail in Section 4.

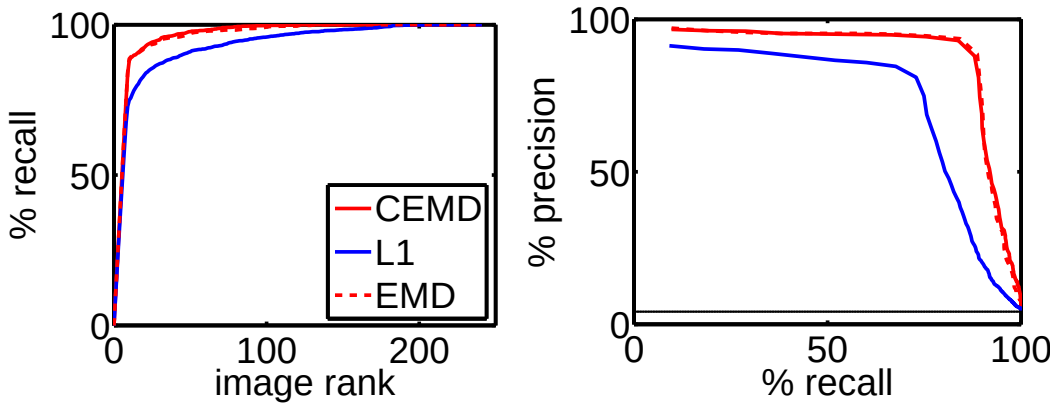
In order to illustrate these phenomena, a small dataset[2] of 22 photographs has been used, shown in Figure 6. Here, we propose to reproduce –in a synthetic way– the color image retrieval experiment presented in section 3.2.2. For each picture of this dataset, synthetic modifications are proceeded in order to simulate two types of perturbations which naturally arise when considering color image retrieval:

- **Gamma** correction with a power factor varying from 0.6 to 2.4 (this operation has been realized on the "Luminance" channel in CIELab color space). An example is shown in Figure 7(a);
- **White balance** correction with a "color temperature" varying from 4400 to $6200^\circ K$ (Example is given in Figure 8(a)).

Fig. 6 22 pictures used for image retrieval test (results are shown in Figures 7 and 8).

Now, applying these modifications to the dataset, we obtain two different databases on which a retrieval is performed using L^1 and CEMD metric (like in section 3.2.2). Results are shown in Figure 7 for gamma correction, and in Figure 8 for color temperature correction.

(a) Example of gamma modification



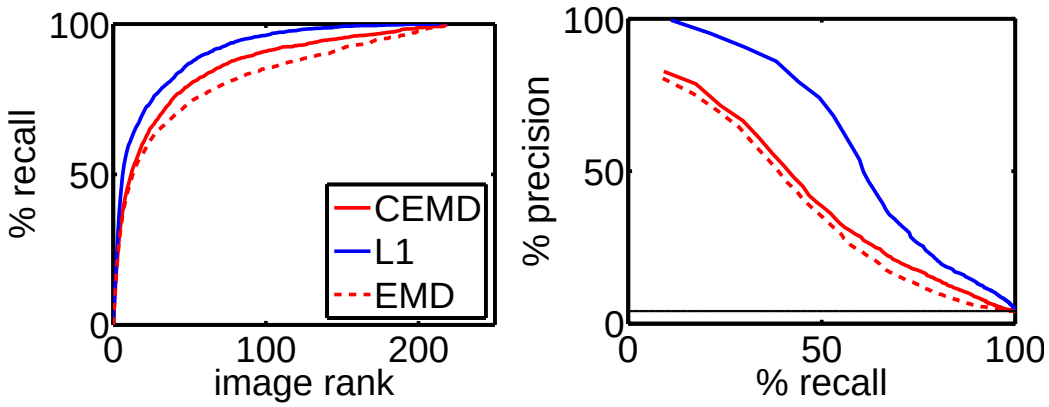
(b) Average recall rate (c) Precision-recall curve

Fig. 7 Retrieval results with data corrupted by contrast modification (gamma correction).

[2] the image dataset is available at the following address: http://perso.telecom-paristech/~rabin/database/

Once again, in the case of gamma correction (Figure 7) CEMD provides in average better retrieval results than L^1 distance. The main reason in such case is that we observe some **intraclass shifts** between histograms, for which cross-bin distances such as CEMD are more robust than bin-to-bin distances.

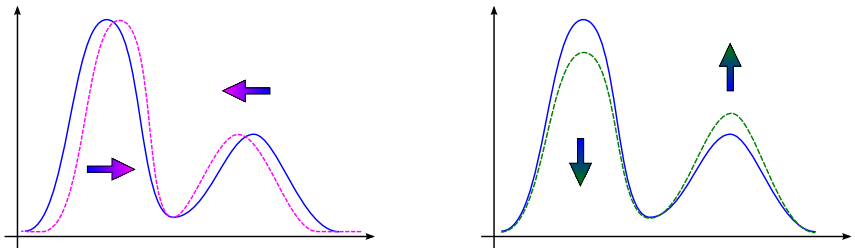
(a) Example of color temperature modification



(b) Average recall rate (c) Precision-recall curve

Fig. 8 Retrieval results with data corrupted by white balance modification (color temperature correction).

Now, in the case of color temperature modification (Figure 8), one observes the following result: L^1 distance provides better retrieval scores than CEMD. An examination of the results has led us to observe that, in such a case, the intraclass variability results this time from **differences of weight** of dominant modes in histograms (see Figure 9(b) for an illustration).



(a) Illustration of histogram shift (b) Illustration of histogram weight variability

Fig. 9 Illustration of the two main classes of perturbations involved in retrieval performances: intraclass shift variability (to the left) and intraclass weight variability (to the right).

In order to understand the implications of these results, a discussion is proposed in the following section.

4 Is it worth using transportation distances to compare histograms ?

Following the last two experiments of the previous section (gamma correction and color balance), this section provides a discussion on the relative advantages of Monge-Kantorovich distances using convex cost functions, those using concave cost functions, and bin-to-bin distances. The discussion is not specific to the circular case and will be made from non-circular synthetic examples.

Writing as before $d(x, y)$ for the geodesic distance on the circle, we consider the following distances:

- the L^1 bin-to-bin distance,
- Monge-Kantorovich distances with concave cost functions:
  - $\mathrm{MK}_{\exp\tau}$ defined from Formula (3) when using the exponential cost function $c(x,y) = 1 - \exp(-\frac{d(x,y)}{\tau})$
  - $\mathrm{MK}_{T\tau}$ defined from Formula (3) when using a thresholded cost function as introduced in [PW08, PW09], that is, $c(x,y) = \min(d(x,y), \tau)$ (see Section 2.3.2).
- Monge-Kantorovich distances with convex cost functions: $\mathrm{MK}_\lambda = (\mathrm{MK}_c)^{\frac{1}{\lambda}}$, with MK_c the quantity defined by Formula (3) when using a cost fonction $c(x,y) = d(x,y)^\lambda$, for $\lambda \geq 1$.

Recall that among these distances, only L^1, MK_λ and MK_{T2}can be computed in linear time. Observe also, following the remarks of Section 2.3.2 on the proximity between $\mathrm{MK}_{T2}$and L^1, that these distances in a sense produce a complete range of alternatives between bin-to-bin distances (such as L^1) and Monge-Kantorovich associated with highly non-convex cost functions (e.g. MK_3). This fact will be quite clear in the following synthetic experiments.

These experiments consist, in order to study the assets of the various distances, to perform retrieval from synthetic histograms (mixture of two Gaussians) in the presence of two types of perturbations: shifts in the positions of bins on the one hand, and variation in the weight of bins on the other hand (see Figure 9). Observe that these two types of perturbations correspond to the ones encountered at the end of Section 3.2.2.

We assume that elements to be compared belong to two classes A and B, and that each element is represented by one N-bins histogram. We model the histograms as the mixture of two Gaussians. Writing $c \in \{A, B\}$ for the class, these two Gaussians have weights p^c and $(1-p^c)$, means μ_1^c and μ_2^c, and standard deviations σ_1^c and σ_2^c (see Figure 10). In the following experiments, parameters are set as follows

- *Histogram construction* Quantization of histograms: $N = 100$ bins; Number of samples for Gaussian mixture data generation: $1,000$ samples in $[0,1]$; Number of histograms per class: $1,000$ histograms.
- *Gaussian mixture parameters* Weights: $p^A = 0.6$ and $p^B = 0.8$; Means: $\mu_2^A = \mu_2^B = 0.2$ and $\mu_2^A = \mu_2^B = 0.7$; Standard-deviations: $\sigma_1^A = \sigma_1^B = \sigma_2^A = \sigma_2^B = 0.05$.

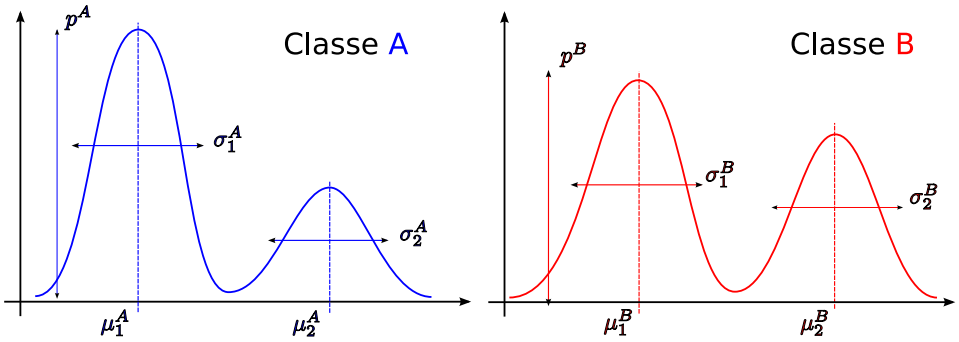
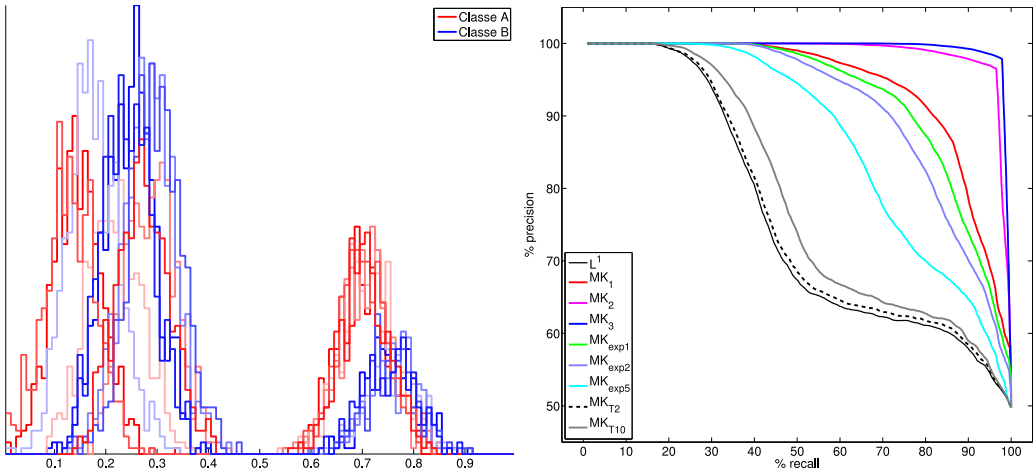


Fig. 10 The two classes A and B are defined as a Gaussian mixture model. For each class, the two Gaussian distributions are defined with 4 parameters (means and standard-deviations), plus a weighting parameter p.

This generative model being chosen, two different kinds of variability can now be simulated to evaluate the robustness of transportation distances depending on the cost function [3] (see Figure 9).

Histogram shift We introduce random shifts in the histogram by modeling the means μ_1^c as random variables. We choose $\mu_1^A = 0.2 + \epsilon_\mu$, where ϵ_μ is uniformly drawn in $[-0.1; 0.1]$. Some of such generated histograms are superposed in Figure 11(a). The precision-recall curves resulting from this two-class retrieval problem are plotted for different metrics in Figure 11(b). One first observes that distances MK_λ, relying on convex cost functions, give the best results, the larger λ the better. Second, it can be seen that transportation distances with concave cost function yields less efficiency. First are distances relying on an exponential cost. Eventually using transportation distances with truncated L^1 distances provides poor results, similar to those obtained with the L^1 distance. This fact is in agreement with the analysis made in Section 2.3.2.

[3] The Earth Mover's Distances with exponential and truncated cost functions have been computed using the code kindly provided by Y. Rubner [Rub].



(a) Histogram shift variability experiment (b) Precision-Recall curve

Fig. 11 Two-class retrieval problem with intraclass shift variability. The effect of the perturbation on histograms is shown in Figure 11(a). The Precision-Recall curves are displayed in Figure 11(b) for several transportation distances: $\mathrm{MK}_{T\tau}$ refers to as the transportation distance with truncated cost function according to the threshold $\tau \in \{2, 10\}$, $\mathrm{MK}_{\exp\tau}$ corresponds to the transportation distance with exponential cost function using parameter $\tau \in \{1, 2, 5\}$, and MK_λ is the Monge-Kantorovich distance with $\lambda \in \{1, 2, 3\}$. In addition is shown the curve obtained with L^1 metric, which is equivalent to MK_{T1} (see § 2.3.1).

Histogram weight variability In the second experience, intraclass weight variability are now simulated by modeling weights as random variables: $p_1^A = 0.6 + \epsilon_p$, where ϵ_p is uniformly drawn from $[-0.1; 0.1]$. Some of such generated histograms are superposed in Figure 12(a). The precision-recall curves resulting from this two-class retrieval problem are plotted for different metric in Figure 12(b). One observes that with this kind of perturbation, transportation distances with L^1 cost function are less robust than the L^1 distance. This time, it can be seen that distances with concave cost function yield better retrieval performances. Using thresholded cost functions again provides results that are very similar to those obtained with the L^1 distance. In the meantime, distances relying on exponential cost functions are still half-way between convex cost functions and thresholded cost functions.

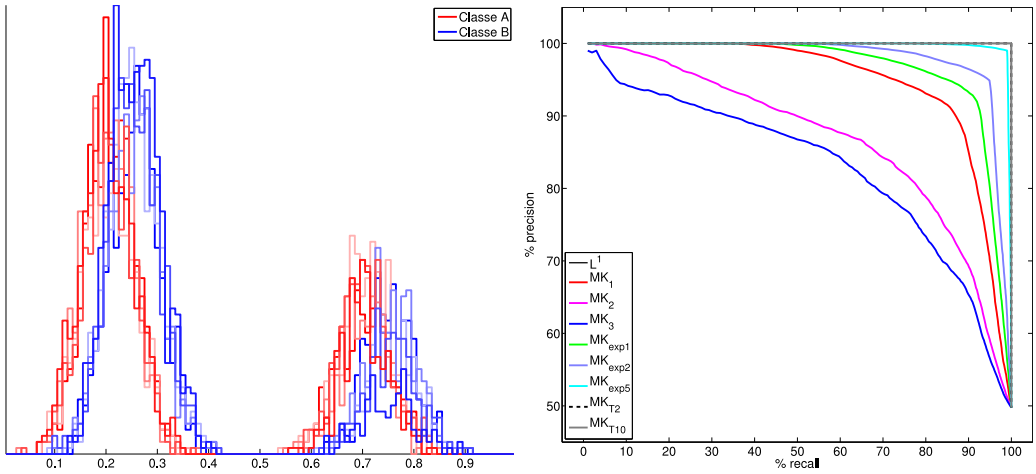
It therefore appears that higher robustness to one type of perturbation yields poorer robustness to the other type. There is a logical tradeoff between robustness to shifts and weight variability. In this context, and given that it may be computed in linear time, the MK_1 distance appears as a good compromise in term of computational cost and robustness to the two kinds of variability considered here.

5 Conclusion

In this paper, the optimal mass transport problem on the circle has been addressed in the case of convex and increasing cost function of the geodesic distance on the circle. We have proposed a new formulation (and a proof) for estimating the corresponding Monge-Kantorovich distances. In the particular case where the cost function is the geodesic distance on the circle, it has been shown that the transportation distance MK_1 between circular histograms (also referred to as CEMD, standing for Circular Earth Mover's Distance) can be deduced by a very simple Formula (12) which is computed in linear time.

Then, several applications in this framework has been studied (hue transfer, local features comparison and color image retrieval), exploiting both the optimal transportation cost between histograms but also the corresponding optimal flow. Other applications could also benefit from the CEMD metric, such as shape recognition based on circular descriptors (see *e.g.* character recognition with orientation histogram [CS02], and curvature based descriptor along closed contour [Mok97, JWR06]).

In the last section, a comparative analysis of transportation distances with different cost functions has been proposed, considering two types of perturbations which arise with histogram representation: mean and weight changes of dominant modes. We have demonstrated that there is a tradeoff between these two phenomena when using either convex or concave cost functions. Eventually, the proposed CEMD metric offers an interesting compromise between these two choices, while being easy to use.



(a) Histogram weight variability experiment (b) Precision-Recall curve

Fig. 12 Two-class retrieval problem with intraclass weight variability. The effect of the perturbation on histograms is shown in Figure 12(a). The Precision-Recall curves are displayed in Figure 11(b), plotted for different transportation distances.

Acknowledgements Delon acknowledges the support of the French Agence Nationale de la Recherche (ANR), under grant BLAN07-2_183172, Optimal transport: Theory and applications to cosmological reconstruction and image processing (OTARIE), and would like to thank J. Salomon and A. Sobolevski for fruitful discussions.

A Appendix: Proof of Theorem 1

This appendix provides a complete proof of Theorem 1 in the case where f and g are discrete distributions (as written in Equation (2)). We first prove this theorem for distributions composed of unitary masses, and conclude thanks to continuity arguments.

A.1 Introduction

Consider two discrete sets of points $\{x_1, \dots x_P\}$ and $\{y_1, \dots y_P\}$ on the unit circle S^1, and the corresponding discrete distributions

$$f = \frac{1}{P}\sum_{k=1}^{P} \delta_{x_k}, \text{ and } g = \frac{1}{P}\sum_{k=1}^{P} \delta_{y_k}, \tag{16}$$

where the notations x_k, y_k are used equally for points on the unit circle or for their coordinates in $[0, 1[$. Let d be the geodesic distance along the circle (given by Equation (5)) and assume that c can be written $c(x, y) = h(d(x, y))$ with h a nonnegative, increasing and convex function. It is well known (this is a consequence of Birkhoff's theorem, see for example the introduction of [Vil03]) that the optimal transportation cost between f and g equals

$$\mathrm{MK}_c(f, g) = \min_{\sigma \in \Sigma_P} W_\sigma^c(f, g), \quad \text{with} \quad W_\sigma^c(f, g) := \frac{1}{P}\sum_k c(x_k, y_{\sigma(k)}) = \frac{1}{P}\sum_k h(d(x_k, y_{\sigma(k)})), \tag{17}$$

where Σ_P is the set of permutations of $\{1, \dots P\}$. In other words, finding the optimal transportation between f and g boils down to find the optimal permutation σ between the points $\{x_k\}$ and $\{y_j\}$.

A.1.1 Paths

If x and y are two different points of S^1, we note $\gamma(x, y)$ **the geodesic path linking** x **and** y **on** S^1 (the path is supposed open: it does not contain x and y). This path is always unique except in the case where

x and y are in opposite positions on the circle. In this case, we choose $\gamma(x, y)$ as the path going from x to y in the trigonometric direction. A path $\gamma(x, y)$ is said to be **positive** if it goes from x to y in the trigonometric direction. If the path goes from x to y in the opposite direction , it is said to be **negative**.

A.1.2 Cumulative distribution functions

The cumulative distribution function of f has been defined in Equation (6). Now, on $[0, 1[$ seen as a the unit circle S^1, no strict order can be defined between points, which means that we can define as many cumulative distribution functions as there are starting points on the circle. If x is a point in $[0, 1[$, the x-cumulative distribution function F_x of f can be defined by choosing x as the reference point on the circle S^1 and by summing the mass in the trigonometric order from this new reference point:

$$\forall y \in \mathbb{R}, \quad F_x(y) = F(x + y) - F(x). \tag{18}$$

An example of a cumulative distribution F and its corresponding x-cumulative distribution F_x on $[0, 1[$ is shown on Figure 13.

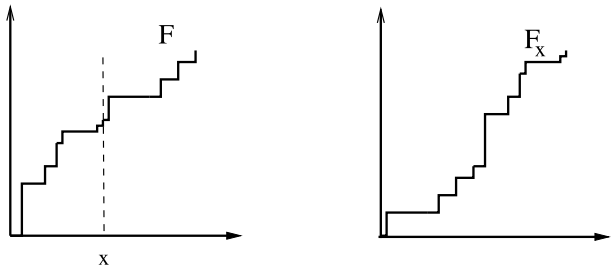


Fig. 13 F on the left and F_x on the right.

A.2 Preliminary results

In the following, we prove that if f and g can be written as in Equation (16), if the points $x_1, \dots x_P$ and $y_1, \dots y_P$ are pairwise different, and if σ is an optimal permutation for (17), there is always a point on the circle which is not contained in any optimal path of σ. This result is proven first for strictly convex functions h and for any optimal permutation σ, then for convex functions h and a well chosen optimal permutation.

Proposition 1 *Assume that h is strictly convex. Let $x_1, \dots x_P$ and $y_1, \dots y_P$ be P points in $[0, 1[$, all pairwise different. Then for each permutation σ of Σ_P which minimizes (17) , there exists $k \in \{1, \dots P\}$ such that for all $l \neq k$, $x_k \notin \gamma(x_l, y_{\sigma(l)})$.*

The proof of this proposition needs the following lemma, which describes some properties of the geodesic paths $\gamma(x_l, y_{\sigma(l)})$ obtained when σ is a minimizer of (17) and h is strictly convex.

Lemma 1 *Assume that h is strictly convex. Let σ be a minimizer of (17) and let $\gamma_l = \gamma(x_l, y_{\sigma(l)})$ and $\gamma_k = \gamma(x_k, y_{\sigma(k)})$ (with $l \neq k$) be two geodesic paths for the assignment defined by σ. Assume also that $x_l \neq x_k$ and $y_{\sigma(l)} \neq y_{\sigma(k)}$. Then, one of the following holds:*

- $\gamma_l \cap \gamma_k = \emptyset$ *;*
- $\gamma_l \cap \gamma_k \neq \emptyset$ *and in this case γ_l and γ_k have the same direction (both positive or both negative) and neither of them is contained in the other.*

Proof Assume that $\gamma_l \cap \gamma_k \neq \emptyset$. If $\gamma_l \cap \gamma_k$ is equal to $\gamma(x_l, x_k)$, then, since h is an increasing function of d, $c(x_l, y_{\sigma(l)}) > c(x_k, y_{\sigma(l)})$ and $c(x_k, y_{\sigma(k)}) > c(x_l, y_{\sigma(k)})$, which contradicts the optimality of σ. The same conclusion holds if $\gamma_l \cap \gamma_k$ is equal to $\gamma(y_{\sigma(l)}, y_{\sigma(k)})$. Moreover, if for example the path γ_l is included in γ_k, then the strict convexity of the function h implies

$$c(x_l, y_{\sigma(l)}) + c(x_k, y_{\sigma(k)}) > c(x_l, y_{\sigma(k)}) + c(x_k, y_{\sigma(l)}),$$

which also contradicts the optimality of σ. Thus, $\gamma_l \cap \gamma_k$ is equal to $\gamma(x_l, y_{\sigma(k)})$ or to $\gamma(x_k, y_{\sigma(l)})$ and it follows that γ_k and γ_l are either both positive or both negative.

Proof of Proposition 1 Let σ be a minimizer of (17). In the following, we will denote by γ_l the geodesic path $\gamma(x_l, y_{\sigma(l)})$. We can assume without loss of generality that the points $x_1, \dots x_P$ are in trigonometric order on the circle.

Assume that for each $l \in \{1, \dots P\}$, there exists $q(l) \neq l$ such that x_l belongs to the open path $\gamma_{q(l)}$. Then, for each l, we have $\gamma_{q(l)} \cap \gamma_l \neq \emptyset$, which means that the geodesic paths $\gamma_{q(l)}$ and γ_l are either both positive or both negative (from lemma 1). Assume for instance that they are both positive and let us show that in this case $x_l \in \gamma_{l-1}$ (with $l-1 = P$ if $l = 0$). If $q(l) = l-1$, there is nothing to prove. If $q(l) \neq l-1$, it means in particular that $x_{q(l)}, x_{l-1}, x_l$ are in trigonometric order on the circle. Since $\gamma_{q(l)}$ is a positive path starting from $x_{q(l)}$ and containing x_l, it follows that $\gamma_{q(l)}$ contains x_{l-1} (recall that the points are assumed to be pairwise different, in particular $x_{l-1} \neq x_{q(l)}$). Thus $\gamma_{l-1} \cap \gamma_{q(l)} \neq \emptyset$, which implies that γ_{l-1} is positive. Now, x_l must be in γ_{l-1}, otherwise we would have $\gamma_{l-1} \subset \gamma_{q(l)}$, which contradicts lemma 1. Thus, if the paths $\gamma_{q(l)}$ and γ_l are both positive, $x_l \in \gamma_{l-1}$.

In the same way, if $\gamma_{q(l)}$ and γ_l are both negative, then $x_l \in \gamma_{l+1}$. In any case, for each $l \in \{1, \dots P\}$, $x_l \in \gamma_{l-1} \cup \gamma_{l+1}$ (with the obvious convention $\gamma_{P+1} = \gamma_1$, $\gamma_0 = \gamma_P$).

Now, suppose that for a given $k \in \{1, \dots P\}$, x_k is in γ_{k-1}. Then, γ_{k-1} and γ_k have the same direction. From lemma 1, it follows that x_{k-1} cannot be contained in γ_k. Since we know that $x_{k-1} \in \gamma_{k-2} \cup \gamma_k$, x_{k-1} must be in γ_{k-2}. Recursively, for each $l \in \{1, \dots P\}$, $x_l \in \gamma_{l-1}$. It follows that for each $l \in \{1, \dots P\}$, $d(x_l, y_{\sigma(l-1)}) < d(x_{l-1}, y_{\sigma(l-1)})$, and since h is increasing

$$\sum_{l=1}^{P} c(x_l, y_{\sigma(l)}) > \sum_{l=1}^{P} c(x_{l+1}, y_{\sigma(l)}), \tag{19}$$

which contradicts the fact that σ is a minimizer of (17). We come to the same conclusion if for a given $k \in \{1, \dots P\}$, x_k is in γ_{k+1} □

The same result can be proven for any convex function h with the difference that it is only satisfied for a good choice of the permutation σ which minimizes (17), and not for all of these permutations. This result can be seen as a limit version of proposition 1.

Corollary 1 *Assume that h is convex. Let $x_1, \dots x_P$ and $y_1, \dots y_P$ be P points in $[0, 1[$. Assume that all these points are pairwise different. Then there exists a permutation σ of Σ_P which minimizes (17) and a point $x_k \in \{x_1, \dots x_P\}$ such that for all $l \neq k$, $x_k \notin \gamma(x_l, y_{\sigma(l)})$.*

Proof We know that for any strictly convex function h, if σ_h minimizes the cost $\sigma \mapsto W_\sigma^c(f, g)$, there exists $k \in \{1, \dots P\}$ such that for all $l \neq k$, $x_k \notin \gamma_l = \gamma(x_l, y_{\sigma_h(l)})$.

Now, assume that h is convex (not strictly). One can always find a sequence (h^n) of increasing and strictly convex functions such that h^n converges pointwise towards h when $n \to \infty$. If σ and the points $x_1, \dots x_P, y_1, \dots y_P$ are fixed, then the finite sum $W_\sigma^n(f, g) := \frac{1}{P} \sum_k h^n(d(x_k, y_{\sigma(k)}))$ tends towards $W_\sigma(f, g) = \frac{1}{P} \sum_k h(d(x_k, y_{\sigma(k)}))$ when $n \to \infty$. Thus, for each $\varepsilon > 0$, there exists an integer N, such that for all $n \geq N$, $|W_\sigma^n(f, g) - W_\sigma(f, g)| \leq \varepsilon$. Since Σ_P is a finite set, we can chose N large enough such that this property holds for every σ in Σ_P. We can also chose N such that $|\min_\sigma W_\sigma(f, g) - \min_\sigma W_\sigma^n(f, g)| \leq \varepsilon$. Now, if $n \geq N$ and if σ^* is an optimal permutation for $W_\sigma^n(f, g)$, it follows that

$$\begin{aligned} |\min_\sigma W_\sigma(f, g) - W_{\sigma^*}(f, g)| &\leq |\min_\sigma W_\sigma(f, g) - \min_\sigma W_\sigma^n(f, g)| \\ &\quad + |W_{\sigma^*}^n(f, g) - W_{\sigma^*}(f, g)| \\ &\leq 2\varepsilon. \end{aligned}$$

Since Σ_P is a finite set, the fact that this distance can be made arbitrarily small implies that when n is large enough, a minimizer σ^* of $W_\sigma^n(f, g)$ is also a minimizer of $W_\sigma(f, g)$. This proves that there exists at least one minimizer σ of $\sigma \mapsto W_\sigma(f, g)$ such that $x_k \notin \gamma(x_l, y_{\sigma_\lambda(l)})$ for some $k \in \{1, \dots, P\}$ and all $l \neq k$.

A.3 Proof of Theorem 1

Proof of Theorem 1

Let us begin with the case where f and g can be written as sums of unitary masses (Equation (16)), and where $x_1, \dots x_P$ and $y_1, \dots y_P$ are pairwise different. Proposition 1 and Corollary 1 show that if the ground cost c can be written $c(x,y) = h(d(x,y))$ with h a positive, convex and increasing function, we can choose some optimal permutation σ for which there is some point x_k which is not contained in any path of σ (recall that paths are defined as open: they do not contain their boundaries). Since all points are supposed pairwise different, the only path meeting all the neighborhoods of x_k is γ_k. It follows that there exists some open set on one side of x_k and not containing x_k which does not cross any path of the optimal permutation σ. The middle x of this open set is not contained in any path of σ. We can thus cut the circle S^1 at x and reduce the transportation problem on the circle to the transportation problem on the real line. The optimal permutation σ is thus given by the sorting of the points (formula (72) in [Vil03]), taking x as the reference point on the circle. This means that when points are pairwise different, we have

$$\mathrm{MK}_c(f,g) = \inf_{x \in S^1} \int_0^1 h(|F_x^{-1} - G_x^{-1}|), \tag{20}$$

where F_x^{-1} and G_x^{-1} are the pseudo-inverses (pseudo-inverses are defined in Section 2.2) of the increasing functions F_x and G_x defined in Equation (6).

Now, observe that F_x and G_x are horizontal translations of $F - F(x)$ and $G - G(x)$ by the same vector x. In consequence,

$$\int_0^1 h(|F_x^{-1} - G_x^{-1}|) = \int_0^1 h(|(F - F(x))^{-1} - (G - G(x))^{-1}|). \tag{21}$$

Since F and G have been defined on $\mathbb{R}$ such that for all y, $F(y+1) = F(y)+1$ and $G(y+1) = G(y)+1$, the bounds of this integral can be replaced by any bounds $(t, t+1)$. It follows that

$$\begin{aligned}\int_0^1 h(|(F - F(x))^{-1} - (G - G(x))^{-1}|) &= \int_{-F(x)}^{1-F(x)} h(|(F - F(x))^{-1} - (G - G(x))^{-1}|) \\ &= \int_0^1 h(|(F)^{-1} - (G + F(x) - G(x))^{-1}|).\end{aligned}$$

Finally,

$$\mathrm{MK}_c(f,g) = \inf_{x \in S1} \int_0 1h(|(F)^{-1} - (G + F(x) - G(x))^{-1}|). \tag{22}$$

In order to conclude, notice that the function $\varphi : \alpha \mapsto \int_0 1h(|(F)^{-1} - (G+\alpha)^{-1}|)$ is continuous ($h : \mathbb{R} \to \mathbb{R}^+$ is continuous since it is convex) and coercive ($\varphi(\alpha) \to +\infty$ when $|\alpha| \to +\infty$). It follows that φ reaches its minimum at a point $\alpha_0 \in \mathbb{R}$. In addition, the fact that F and G are piecewise constant implies that φ is piecewise affine, with discontinuities of φ' at points $F(x) - G(x)$. Thus,

$$\mathrm{MK}_c(f,g) = \inf_{\alpha \in \mathbb{R}} \int_0 1h(|(F)^{-1} - (G+\alpha)^{-1}|). \tag{23}$$

The previous result can be generalized to the case where the points x_i, y_j may coincide just by remarking that both quantities in Equation (23) are continuous in the positions of these points. In consequence, the result holds for distributions with rational masses.

In order to generalize the result to any couple of discrete probability distributions, observe that the right term in Equation (23) is continuous in the values of the masses $f[i]$ and $g[j]$. As for the continuity of $\mathrm{MK}_c(f,g)$, assume that a mass ε of the distribution f is transferred from the point x_{i_0} to the point x_{i_1} in f, and let us call the new distribution f^ε. If (α) is an optimal transport plan between f and g, let j_0 be an index such that $\alpha_{i_0,j_0} \geq \varepsilon$. A transport plan (α') between f^ε and g can be defined as

- $\alpha'_{i_0,j_0} = \alpha_{i_0,j_0} - \varepsilon$,
- $\alpha'_{i_1,j_0} = \alpha_{i_1,j_0} + \varepsilon$,
- $\alpha'_{i,j} = \alpha_{i,j}$ for $(i,j) \neq (i_0,j_0), (i_1,j_0)$.

The corresponding transportation cost between f^ε and g is then lower than $\mathrm{MK}_c(f,g) + \varepsilon h(\frac{1}{2})$, which implies that $\mathrm{MK}_c(f^\varepsilon, g) \leq \mathrm{MK}_c(f,g) + \varepsilon h(\frac{1}{2})$. Conversely, we can show that $\mathrm{MK}_c(f,g) \leq \mathrm{MK}_c(f^\varepsilon, g) + \varepsilon h(\frac{1}{2})$. □

References

ACB+03. L. Ambrosio, L.A. Caffarelli, Y. Brenier, G. Buttazzo, and C. Villani. *Optimal Transportation and Applications*, volume 1813 of *Lecture Notes in Mathematics*. Springer, Berlin / Heidelberg, mathematics and statistics edition, 2003.

BDM09. R. Burkard, M. Dell'Amico, and S. Martello. *Assignment Problems*. SIAM, 2009.

BMP02. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.

CM95. CA Cabrelli and UM Molter. The Kantorovich metric for probability measures on the circle. *Journal of Computational and Applied Mathematics*, 57(3):345–361, 1995.

CM98. CA Cabrelli and UM Molter. A linear time algorithm for a matching problem on the circle. *Information processing letters*, 66(3):161–164, 1998.

CS02. Sung-Hyuk Cha and Sargur N. Srihari. On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355–1370, June 2002.

Cul06. M. J. P. Cullen. *A Mathematical Theory of Large-Scale Atmospheric-Ocean Flow*. Springer, imperial college press edition, mars 2006.

DSS10. J. Delon, J. Salomon, and A. Sobolevskii. Fast transport optimization for Monge costs on the circle. *SIAM Journal on Applied Mathematics*, 2010. to appear.

Dvi02. Guy Dvir. Context-based image modelling. In *ICPR '02: Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 4*, page 40162. IEEE Computer Society, 2002.

FMMS02. Uriel Frisch, Sabino Matarrese, Roya Mohayaee, and Andrei Sobolevski. A reconstruction of the initial conditions of the universe by optimal mass transportation. *Nature*, 2002.

GD04. K. Grauman and T.J. Darrell. Fast contour matching using approximate earth mover's distance. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, pages I: 220–227, 2004.

GDR00. H. Greenspan, G. Dvir, and Y. Rubner. Region correspondence for image matching via emd flow. In *CBAIVL '00: Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'00)*, page 27, Washington, DC, USA, 2000. IEEE Computer Society.

GM96. W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

Gur90. C. Gurwitz. Weighted median algorithms for L1 approximation. *BIT Numerical Mathematics*, 30(2):301–310, 1990.

HGS08. T. Hurtut, Y. Gousseau, and F. Schmitt. Adaptive image retrieval based on the spatial organization of colors. *Comput. Vis. Image Underst.*, 112(2):101–113, 2008.

HZTA04. Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *Int. J. Comput. Vision*, 60(3):225–240, 2004.

IT03. P. Indyk and N. Thaper. Fast image retrieval via embeddings. In *3rd International Workshop on Statistical and Computational Theories of Vision*, Nice, France, 2003.

JWR06. A.C. Jalba, M.H.F. Wilkinson, and J.B.T.M. Roerdink. Shape representation and recognition through morphological curvature scale spaces. 15(2):331–341, February 2006.

Kan42. L. Kantorovich. On the transfer of masses (in russian). 37(2):227–229, 1942.

LCL04. Qin Lv, Moses Charikar, and Kai Li. Image similarity search with compact data structures. In *CIKM '04: Proceedings of the thirteenth ACM international conference on Information and knowledge management*, pages 208–217, New York, NY, USA, 2004. ACM.

LO07. Haibin Ling and Kazunori Okada. An efficient Earth Mover's distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):840–853, may 2007.

Low04. David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

LZLM05. Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. Region-based image retrieval with high-level semantic color names. In *MMM '05: Proceedings of the 11th International Multimedia Modelling Conference*, pages 180–187, Washington, DC, USA, 2005. IEEE Computer Society.

McC95. R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

McC99. R.J. McCann. Exact solutions to the transportation problem on the line. *Proceedings: Mathematical, Physical and Engineering Sciences*, pages 1341–1380, 1999.

Mok97. Farzin Mokhtarian. Silhouette-based occluded object recognition through curvature scale space. *Mach. Vision Appl.*, 10(3):87–97, 1997.

Mon81. G. Monge. *Mémoire sur la théorie des déblais et des remblais*. Histoire de l'Académie Royale des Sciences, 1781.

PKD07. F. Pitié, A. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, February 2007.

PW08. O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *ECCV08*, 2008.

PW09. Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *ICCV*, 2009.

RDG08. Julien Rabin, Julie Delon, and Yann Gousseau. Circular Earth Mover's Distance for the comparison of local features. In *Proc. ICPR*. IEEE Computer Society, 2008.

RDG09. Julien Rabin, Julie Delon, and Yann Gousseau. A statistical approach to the matching of local features. *SIAM Journal on Imaging Sciences*, 2009.

RT01. Mark A. Ruzon and Carlo Tomasi. Edge, junction, and corner detection using color distributions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1281–1295, 2001.

RTG00. Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The Earth Mover's distance as a metric for image retrieval. *Int. J. Comput. Vision*, 40(2):99–121, 2000.

Rub. Yossi Rubner. Source code for EMD: http://robotics.stanford.edu/~rubner/.

SJ08. S. Shirdhonkar and D.W. Jacobs. Approximate earth mover's distance in linear time. In *CVPR08*, pages 1–8, 2008.

Vil03. C. Villani. *Topics in optimal transportation*. American Math. Soc., 2003.

Vil08. C. Villani. *Optimal transport: old and new*. Springer Verlag, 2008.

WPMK86. M. Werman, S. Peleg, R. Melter, and TY Kong. Bipartite graph matching for points on a line or a circle. *Journal of Algorithms*, 7(2):277–284, 1986.

ZWG06. Qing-Fang Zheng, Wei-Qiang Wang, and Wen Gao. Effective and efficient object-based image retrieval using visual phrases. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 77–80, New York, NY, USA, 2006. ACM.

ZYHT07. L. Zhu, Y. Yang, S. Haker, and A. Tannenbaum. An image morphing technique based on optimal mass preserving mapping. *IEEE Transactions on Image Processing*, 16(6):1481–1495, 2007.